



ANÁLISIS DEL VIENTO EN EL VALLE DEL RÍO NEGRO MEDIANTE MAPAS AUTO ORGANIZADOS Y ALGORITMOS DE INDUCCIÓN

M.G. COGLIATI¹, P. BRITOS² y R. GARCÍA-MARTÍNEZ²

¹Departamento de Geografía. Facultad de Humanidades. Universidad Nacional del Comahue.
e-mail: cogliati@uncoma.edu.ar

²Centro de Ingeniería del Software. Escuela de Graduados. Instituto de Tecnología de Buenos Aires
Laboratorio de Sistemas inteligentes. Facultad de Ingeniería. Universidad de Buenos Aires
e-mail: {pbritos,rgm}@itba.edu.ar

RESUMEN: El presente trabajo incluye el análisis de series temporales de dirección e intensidad de viento en la estación Allen en Río Negro obtenidas durante el desarrollo del Proyecto MECIN, utilizando procedimientos de redes neuronales a partir de la integración de mapas auto organizados y algoritmos de inducción. Se efectuó el análisis de reglas de comportamiento. Se detectaron 7 grupos, y 38 reglas, de las que se seleccionaron 9 que cumplían simultáneamente con las siguientes consideraciones: eran soportadas por más de 25 datos y nivel de confianza superior a 0.84. Esto incluyó el 79% del total de los datos. Los grupos separaron patrones asociables al comportamiento esperable según análisis previos realizados a las series utilizando métodos tradicionales. La inclusión de un número más grande de variables como tiempo y fecha produjo un número mucho mayor de reglas, sin definir intervalos precisos, con pautas confusas y separando tanto pautas obvias como útiles.

Palabras clave: viento, redes neuronales, mapas auto-organizados, algoritmos de inducción.

INTRODUCCIÓN

Las leyes clásicas del movimiento de los fluidos gobiernan el estado de la atmósfera. Los estados atmosféricos exhiben correlaciones en varias escalas espaciales y temporales. El diagnóstico de tales estados procura capturar la dinámica de varias variables atmosféricas (como la temperatura y la presión) y explicar que procesos físicos influyen en el comportamiento. Así un sistema meteorológico se puede pensar como un sistema complejo cuyos componentes interactúan en varias escalas espaciales y temporales. Los enfoques estadísticos a la predicción del tiempo incluyen distintos enfoques (Wilks, 1995; Santhanam y Patra, 2001). Los sistemas inteligentes aparecen como alternativas útiles a las técnicas de modelado estadístico tradicional en muchas disciplinas científicas (Haykin, 1994). Revisando las aplicaciones de redes neuronales en las ciencias atmosféricas, Gardner y Dorling (1998) concluyen que las redes neuronales dan generalmente resultados tan buenos o mejores que los métodos lineales. Hasta ahora, se ha enfocado la atención en combinar los métodos lineales con redes neuronales u otros tipos de sistemas inteligentes para aumentar el poder de este último. Esta tendencia continúa hoy con enfoques más nuevos basados en algoritmos de aprendizaje (Hsieh y Tan, 1998; Monahan, 2000). El término minería inteligente de datos [Evangelos & Han, 1996], es la aplicación de los métodos de aprendizaje automático [Michalski et al., 1983] al proceso no trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y comprensible, a partir de grandes conjuntos de datos, con el objeto de predecir tendencias automatizadas y comportamientos y para describir de forma automatizada modelos previamente desconocidos, (Perichinsky & García-Martínez, 2000;) implicando el uso de técnicas y herramientas de aprendizaje automático (machine learning).

Este análisis se plantea como una alternativa al análisis lineal de series cronológicas meteorológicas. El presente trabajo presenta el análisis de la dirección del viento medio, viento medio y máximo cada 15 minutos en condiciones de terreno complejo. El análisis de series de viento involucran grandes cantidades de datos y la minería de datos podría ser una alternativa a métodos tradicionales estadísticos para encontrar los grupos con señales estables.

DATOS Y METODOLOGÍA

Este estudio presenta un análisis de la dirección e intensidad del viento en una estación meteorológica en Allen (AL) (Río Negro) (39° 01' 05" S, 67° 50' 14" W a 252 m snm) (ver figura 1).

Los datos originales consistieron en un conjunto de datos de temperatura del aire, velocidad y dirección de viento, y presión, obtenidos cada quince minutos desde el 1/10/94 al 17/10/94 en Allen, Río Negro.

La estación meteorológica se localizó en el Valle Superior del río Negro (AVRN). El valle es ancho y poco profundo con bordes en forma de escalón y tiene una orientación de WNW a ESE en el área del estudio. Los datos de la estación

meteorológica se obtuvieron durante el experimento MECIN (MEDiciones de la Capa de Inversión Nocturna) (Flores et al, 1996).

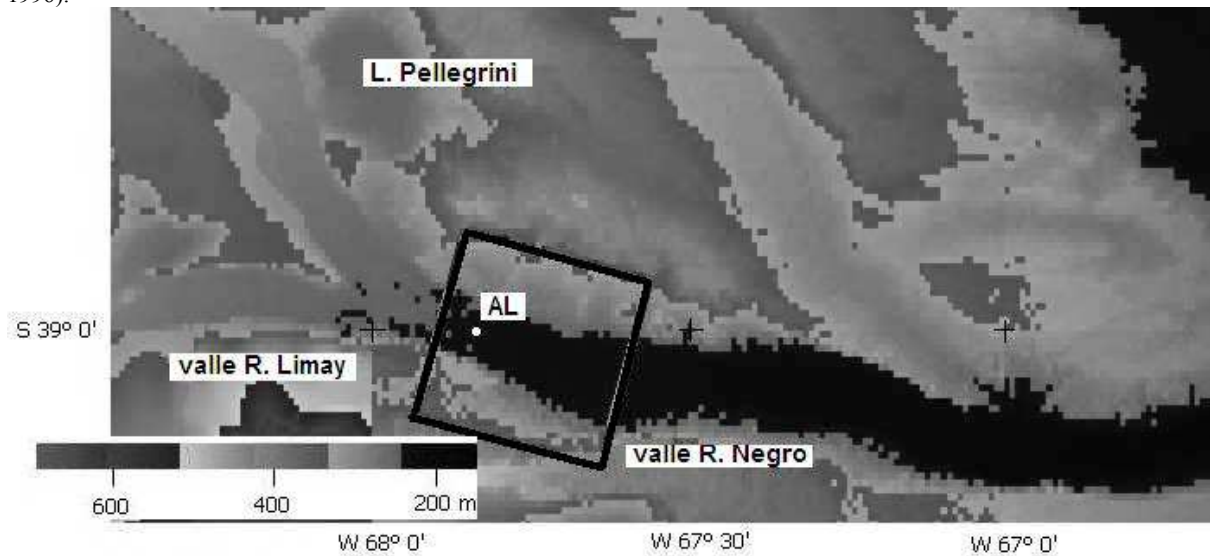


Figura 1: Ubicación geográfica y topografía del Alto Valle del Río Negro y localización de la estación. AL: estación Allen.

El presente trabajo se llevó a cabo por medio del análisis de reglas de comportamiento construidas a partir de la integración de mapas auto organizados (SOM, self organized maps) y algoritmos de inducción (TDIDT, top down induction of decision trees) (Quinlan, 1993) para una serie de dirección e intensidad de viento obtenida para casos particulares en que involucraban intensidades bajas de viento.

Las observaciones fueron hechas bajo condiciones de tiempo de escala sinóptica favorables para la ocurrencia de heladas radiativas (viento en calma y cielo despejado) o inversión nocturna de la temperatura en la capa cercana a la superficie. Estos datos fueron elegidos debido a que se contaba con análisis tradicionales de los mismos en trabajos anteriores (Cogliati, 2001).

Las series temporales de viento y distintas variables meteorológicas consisten a menudo de gran cantidad de observaciones, implicando el análisis de gran cantidad de datos.

Los conjuntos de datos voluminosos, pueden ser analizados desde el punto de vista de la ocurrencia de las formas particulares que distinguen fenómenos específicos. Es entonces posible ver las variables meteorológicas como fuentes de señales espacio-temporales. La información de estas señales puede ser extraída utilizando técnicas de minería de datos. La variación en las variables meteorológicas se puede ver como una mezcla de varias señales espacio-temporales independientes con distinta intensidad. El propósito de este artículo es investigar si existe un conjunto cualquiera de pautas fijas temporales relacionadas a los fenómenos observados del tiempo obtenibles a partir del proceso ilustrado en la figura 2.

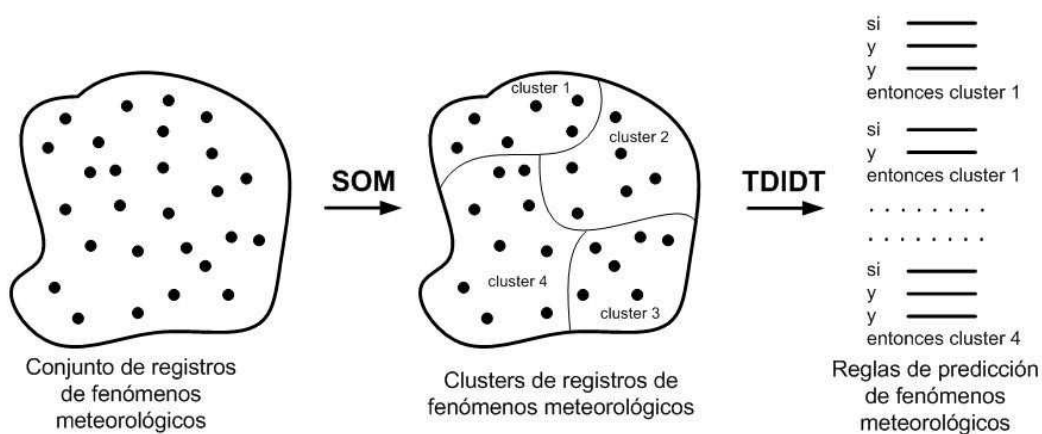


Figura 2: Procesos para establecer comportamientos temporales estables relacionados con fenómenos meteorológicos observados (cluster=grupo).

Nuestra conjetura es que existen patrones, estables e independientes, cuya mezcla determina el comportamiento de las variables; y estos patrones podrían ser extraídas por el análisis de redes neuronales de los datos, vistos como señales temporales. Para tal fin, la serie temporal fue agrupada utilizando redes neuronales auto organizadas o mapas de Kohonen (SOM) (Kohonen, 2001) mediante la herramienta NCLUST y las reglas que describen cada grupo obtenido se construyeron

aplicando algoritmos de inducción (TDIDT) (Quinlan, 1993) a cada grupo. El proceso utilizado se describe en la Figura 1 y determina 7 grupos como se detalla en la Tabla 1.

Grupo	A	B	C	D	E	F	G
Cantidad de datos	342	73	364	80	133	5	206

Tabla 1: Grupos determinados y cantidad de datos por grupo.

RESULTADOS Y DISCUSIÓN

Características del viento en Allen en el período analizado:

En el valle del Río Negro que corre en la dirección WNW – ESE (ver Figura 1), la dirección del viento presentó durante el período entre el 1 y el 17 de octubre de 1994 una gran variabilidad tanto durante el día como durante la noche.

La estación meteorológica utilizada se encontraba ubicada en el interior del valle, en zona de chacras, bajo la influencia de múltiples obstáculos. Tanto para el período analizado en el presente trabajo como en períodos analizados por Cogliati y Mazzeo, (2006), el viento presentó en el período un alto porcentaje de calmas que reflejaría la influencia de la rugosidad del terreno en el interior del valle (49.1% durante la noche y 19.5% durante el día).

Análisis anteriores (Cogliati, 2001; Cogliati y Mazzeo, 2006) enfocan el estudio a partir de los mecanismos que modifican el viento en los valles. Generalmente, en zonas de valles, se reconocen dos clasificaciones de vientos diurnos:

- *vientos de pendiente* que son paralelos a la inclinación de las pendientes y pueden observarse cuesta arriba y cuesta abajo. Estos vientos, son producidos por fuerzas de empuje, inducidas por diferencias de temperatura entre el aire adyacente a la pendiente y el aire ambiental, fuera de la capa límite influenciada por la pendiente (Whiteman, 1990) y

- *vientos de valle*: estos vientos, en general, tienen direcciones valle arriba durante el día y valle abajo durante la noche (Whiteman, 1990). Los vientos debidos exclusivamente a efectos térmicos son difíciles de estudiar, debido a que son influenciados por otros sistemas de viento que se desarrollan en distintas escalas, por gradientes regionales de presión superpuestos a la topografía, por efectos mecánicos inducidos por la topografía así como por sistemas de viento de escalas mayores.

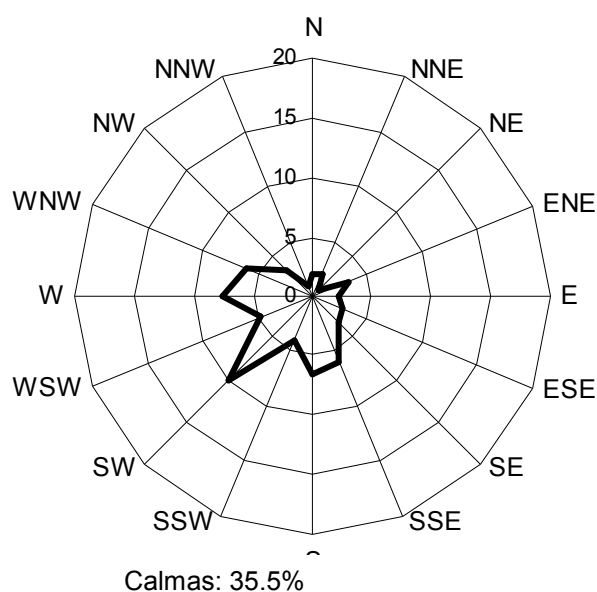


Figura 3: Frecuencia relativa porcentual de la dirección del viento en el período 1 al 17/10/94 en Allen.

Los vientos de valles, soplan paralelos al eje longitudinal del valle. Estos vientos, son producidos por gradientes horizontales de presión que se desarrollan como resultado de las diferencias de temperatura que se forman a lo largo del eje del valle o por diferencias de temperatura entre el aire dentro del valle y el aire a la misma altura sobre la meseta adyacente.

En Allen no se notó el predominio de una dirección particular más notorio en otras estaciones en el valle (Cogliati y Mazzeo 2006) aunque las mayores frecuencias se detectan con viento del sector W. La frecuencia relativa porcentual de la dirección del viento presentó en el período máximos en las direcciones asociadas al sector oeste (22.6%) (SW: 10.0 %, W: 7.6 %, WNW: 6.0 %) y S (6.5 %) (ver Figura 3). Cabe mencionar que en las dos muestras analizadas un porcentaje importante se debe a las calmas. En ambos períodos considerados el porcentaje de calmas fue importante (49.1% entre el 1 y el 17 de octubre de 1994 y 82.4 % entre el 1 y el 5 de septiembre de 1993 durante la noche). Se presentó un mayor porcentaje de casos del WNW (valle abajo) durante el día y se observó un aumento de las direcciones transversales al valle durante la noche que podrían asociarse a este efecto. (Cogliati y Mazzeo, 2006).

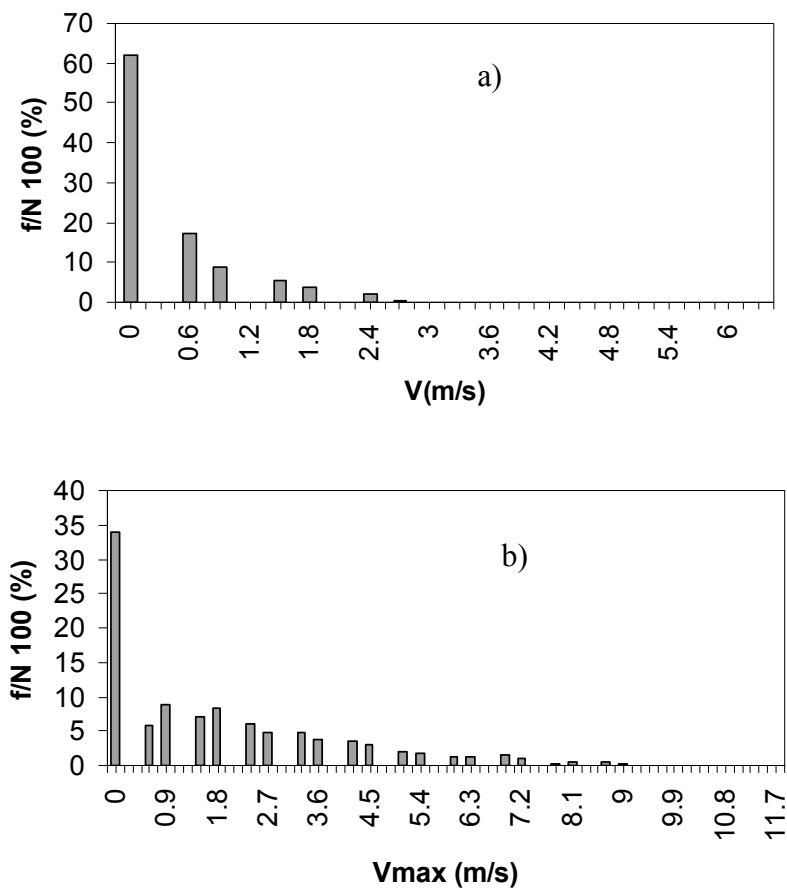


Figura 4: Frecuencia relativa porcentual ($f/N/100\%$, f : frecuencia porcentual, N : total de datos) de intensidad media de viento (m/s) en Allen entre el 1 y el 17/10/94 (a) y Frecuencia relativa porcentual de intensidad máxima de viento (m/s) en Allen entre el 1 y el 17/10/94 (b)

En el período nocturno, los datos reflejaron una canalización forzada (Whiteman, 1990) y durante el día se presentan características de un flujo mucho menos canalizado y con importante el transporte de cantidad de movimiento desde el viento en escala sinóptica (Cogliati y Mazzeo, 2006). Durante la noche se presentaron además casos en la dirección NNE con intensidad de viento muy leve que podrían asociarse a vientos de drenaje desde la barda.

En la Figura 4 se presenta la distribución de frecuencias de intensidad media y máxima del viento, donde puede observarse que los mayores valores de intensidad se presentaron con menor frecuencia mientras que los valores cercanos a la calma presentan la mayor cantidad de datos tanto en la intensidad media como la máxima, que alcanzó en el período 9 m/s.

Análisis de los datos utilizando mapas autoorganizados y algoritmos de inducción

El primer análisis aplicando mapas auto organizados a registros de velocidad media y máxima y dirección de viento, temperatura del aire y presión atmosférica; determinó 7 grupos. La temperatura del aire incluye la variación periódica diaria, que fue incluida para explorar la relación con variaciones diurnas en las direcciones de viento.

El análisis de reglas de comportamiento, determinó 38 reglas, algunas de las cuales incluían una pequeña cantidad de datos. A partir de esto se efectuó una selección de las mismas teniendo en cuenta pautas preestablecidas. Se seleccionaron las reglas que incluían más de 25 datos y alto nivel de confianza (> 0.84). Estas pautas incluyen el 79% de los datos. La tabla 1 presenta las reglas encontradas consignando a que grupo pertenecen, la cantidad de datos soportados y el grado de confianza de las estimaciones, es decir, la probabilidad de ocurrencia de dicha regla.

El grupo A incluyó los valores con presión atmosférica más alta, con viento en calma o débil y temperatura del aire menor a $16.5\text{ }^{\circ}\text{C}$, que se asociarían a los casos dentro de los que se clasifican las heladas radiativas. El grupo B incluyó casos de alta presión y dirección del viento entre el NNE y E. Dentro del grupo C se seleccionó un mayor espectro de direcciones de viento incluyendo el W.

Los casos con presión atmosférica más baja e intensidad del viento por debajo de 0.2 m/s se incluyeron en el Grupo E y las mayores intensidades de viento estuvieron representadas en el grupo G. El grupo F se excluyó debido a que representaba un número muy pequeño de datos.

Grupo	Reglas	Cantidad de Datos	Confianza
A	SI P \geq 981.3 hPa y viento en calma, Vmed $<$ 0.2 m/s y T $<$ 16.5 °C	324	1.00
B	SI P \geq 986.95 hPa y viento dirección NNE a E	42	1.00
C	SI P \geq 986.9 hPa Viento dirección del SE al N por el W y T $<$ 19.1 °C	225	1.00
C	SI P [983.6, 986.9) hPa y dirección SSE a N y Vmax $<$ 3.3 m/s	76	0.84
E	SI P $<$ 981.26 hPa y Vmed $<$ 0.20	55	0.96
E	SI P [981.3, 989.0) hPa y Vmed $<$ 0.2 m/s y Vmax $<$ 1.1 m/s y T \geq 16.5 °C	25	1.00
E	SI P \geq 986.9 hPa Dirección NNE a E y Vmed $<$ 0.2 m/s	33	1.00
G	P $<$ 983.6 hPa y dirección S a NW y Vmax $<$ 3.3 m/s	36	0.89
G	P $<$ 986.5 hPa y S a N por W y Vmax \geq 3.3 m/s	138	0.97

Tabla 2: reglas obtenidas aplicando TDIDT a datos meteorológicos de Allen, Río Negro entre el 1 y 17/10/94 con observaciones cada 15 min.

Cabe mencionar que al realizar los cálculos con la inclusión de fecha y hora como características, se aumentó considerablemente el número de reglas que incluían muy pocos datos, por lo que considerando que las mediciones se efectuaron cada 15 min una pequeña cantidad de datos incluiría periodos muy cortos de tiempo que no podrían asociarse a procesos físicos.

La Figura 4 presenta el diagrama de dispersión de los registros de velocidad máxima vs la hora del día en que ocurrieron, puede notarse una mayor cantidad de casos de intensidades altas en horarios posteriores a la mayor insolación que se asociarían al aumento de la turbulencia del aire, estos casos se incluyen principalmente en el grupo B y G.

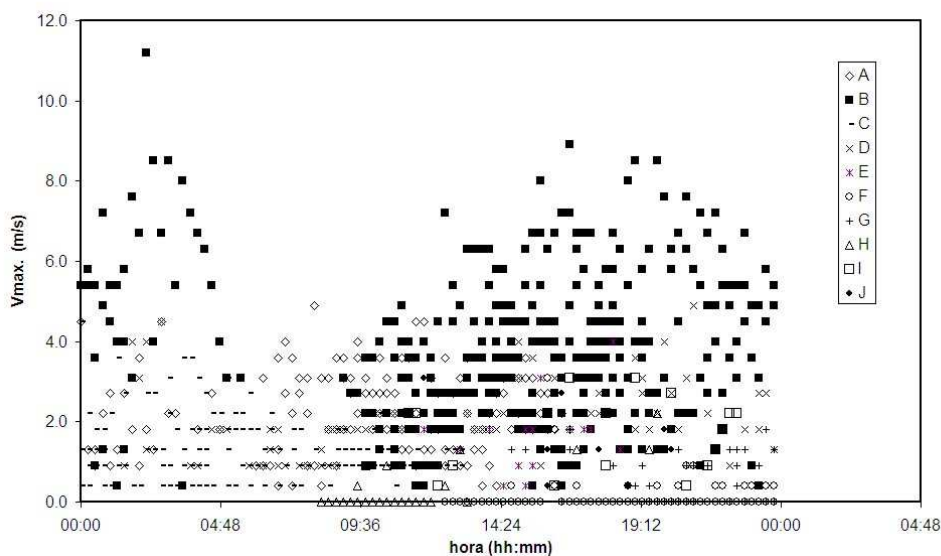


Figura 4. Diagrama de dispersión de los diferentes grupos de datos del viento máximo en Allen asociados a la hora de ocurrencia del episodio. (Río Negro, Argentina) desde 13/10/94 al 17/10/94.

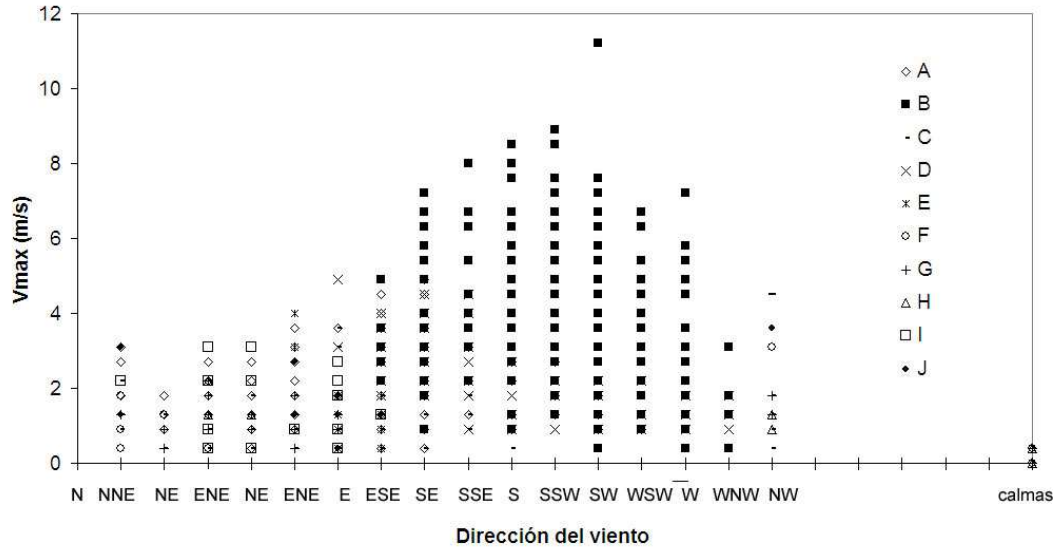


Figura 5: Diagrama de dispersión de la velocidad máxima y la dirección del viento agrupados utilizando SOM, para los datos del 1 al 17 de octubre de 1994 en Allen.

La Figura 5 presenta el diagrama de dispersión de la velocidad máxima y la dirección del viento señalando en cada caso el grupo al cual pertenecen los datos, se pueden observar patrones de distribución de los datos en sectores determinados de dirección del viento. El grupo B se presenta principalmente para velocidades máximas asociadas a direcciones del sector S. Mientras que el Grupo A se presenta en direcciones del sector E. El viento máximo presenta mayores valores del sector Sur coincidiendo con el análisis tradicional, incluyendo datos en los Grupos B y D.

CONCLUSIONES

El comportamiento de las variables temporales puede ser considerada como una mezcla de varias señales espacio-temporales independientes ocurriendo con diferente intensidad que pueden estudiarse a partir del análisis de mapas autoorganizados y algoritmos de inducción.

El análisis de series de viento involucran grandes cantidades de datos y la minería de datos podría ser una alternativa a métodos tradicionales estadísticos para encontrar los grupos con señales estables.

Se realizó un estudio exploratorio de series cronológicas de variables meteorológicas (temperatura, presión, intensidad y dirección de viento) en Allen, en el Alto Valle del Río Negro, utilizando redes neuronales auto-organizadas (SOM) y se detectaron reglas de comportamiento utilizando algoritmos de inducción (TDIDT). Las reglas obtenidas representan características de la dirección e intensidad del viento, asociados al comportamiento de la temperatura del aire y la presión atmosférica.

Se detectaron 7 grupos, y 38 reglas, de las que se seleccionaron 9 reglas que cumplían simultáneamente con las siguientes consideraciones: soportadas por más de 25 datos y nivel de confianza superior a 0.84. Esto incluyó el 79% del total de los datos.

Los grupos separaron calmas y viento débil, patrones de presión atmosférica y su asociación con la dirección e intensidad del viento y patrones asociables al comportamiento esperable según análisis previos realizados a las series utilizando métodos tradicionales (Cogliati, 2001); relaciones nuevas encontradas deberán ser estudiadas en profundidad. La inclusión de un número más grande de variables como tiempo y fecha produjo un número mucho mayor de reglas, sin definir intervalos precisos, con pautas confusas y separando tanto pautas obvias como útiles. En este hecho debería profundizarse el estudio, debido a que la cronología de la ocurrencia de los fenómenos meteorológicos es de interés para el análisis de diversas situaciones atmosféricas.

Cabe mencionar además que algunas de las características de los flujos con comportamientos diurnos podrían detectarse incorporando la información de día y noche de manera adecuada, actualmente se está trabajando en ese sentido.

Agradecimientos: Los autores quieren agradecer al LISAMET y al Ing. Jorge Lässig por proporcionar los datos meteorológicos obtenidos en el experimento de campo de MECIN utilizados en este estudio.

REFERENCIAS BIBLIOGRAFICAS

- Cogliati, M.G. 2001. *Estudio térmico y del flujo del aire en septiembre y octubre en los valles de los ríos Limay, Neuquén y Negro*. Tesis doctoral. Universidad de Buenos Aires.
- Cogliati, M.G., Mazzeo, N.A (2006) Air flow analysis in the upper Río Negro Valley (Argentina) Atmos. Res. 80: 263-279.
- Evangelos, S., Han, J. 1996. *Proceedings of the Second International Conference on Knowledge Discovery y Data Mining*. Portly, EE.UU.
- Flores, A. ; Lässig, J. ; Cogliati, M. ; Palese, C., Bastanski, M. 1996. *Mediciones de la Capa de Inversión Nocturna en los valles de los ríos Limay, Neuquén y Negro*. Anales VII Congreso Argentino de Meteorología. VII Congreso Latinoamericano e Ibérico de Meteorología. Bs. As.
- Gardner, M., Dorling, S. 1998. *Artificial neuronal networks (the multilayer perceptron) – a review of applications in the atmospheric sciences*. Atmospheric Environment 32: 2627-2636
- Haykin, S., 1994. *Neuronal networks: A comprehensive foundation*. Prentice-Hall, Englewood Cliffs, NJ.
- Hsieh, W. , y Tang, B. 1998. *Applying neuronal network models to prediction y data analysis in meteorology y oceanography*. Bulletin of American Meteorological Society 79: 1855-1870.
- Kohonen, T. 2001. *Self-Organizing Maps*. Springer Series in Information Sciences, Vol. 30, Springer, Berlin.
- Michalski, R., Carbonell, J., Mitchell, T. 1983. *Machine learning I: An AI Approach*. Morgan Kaufmann, Los Altos, CA.
- Monahan, A. 2000. *Nonlinear principal component analysis by neuronal networks: Theory y applications to the Lorenz system*. Journal of Climate 13: 821-835.
- Perichinsky, G., García-Martínez, R. 2000. *A Data Mining Approach to Computational Taxonomy*. Proceedings Argentine Computer Science Researchers Worksop: 107-110.
- Quinlan, R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers. San Mateo California.
- Santhanam M., y Patra, P. 2001. *Statistics of atmospheric correlations*. Physical Review E 64: 016102-1-1-7.
- Whiteman, C.D. (1990) Observation of thermally developed wind systems in mountainous terrain. AMS Meteorological Monographs. N 45, 23. Ed. Blumen, W. 5-42.
- Wilks, D. 1995. *Statistical methods in Atmospheric Sciences*. Academic Press, London.

ABSTRACT

The present work includes the analysis of time series of wind speed and wind direction in Allen automatic weather station in Negro River valley, obtained during the Mecin Project, using neuronal networks procedures. The behavior rules were obtained from the analysis of self organized maps and algorithms of induction. The methodology detected 7 groups, and 38 rules, of which, 9 rules were selected because they simultaneously follow: inclusion of more than 25 data and its confidence level was above 0.84. This facts included the 79% of the total data. The behavior of separated clusters matches prior analysis carried out using traditional methods. The inclusion of a larger number of variables as time and date produced a greater number of rules, without clear defining intervals, with confused guidelines and separating so much obvious patterns like useful ones.

Keywords: wind, neuronal networks, self organized maps, top down induction of decision trees.